Ümit Özgüner (S'72–M'75–M'80) received the M.Sc. degree from the Istanbul Technical University, Istanbul, Turkey, in 1971 and the Ph.D. degree from the University of Illinois, Champaign-Urbana, in 1975, both in electrical engineering.

He was a Teaching Assistant at the Department of Electrical Engineering and a Research Assistant at the Coordinated Science Laboratory, while at the University of Illinois. From May 1975 to May 1976 he had a Post-Doctoral Research Appointment at the IBM Research Center, Yorktown Heights, NY. He was then an Assistant Professor of Electrical Engineering at the Istanbul Technical University, where he was given the title of "Dozent" in April 1980. During the year 1979–1980 he was on leave at the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada. He joined the faculty of the Ohio State University, Columbus, Department of Electrical Engineering, in January 1981. His interests are in linear systems, large-scale decentralized systems control, and microprocessors for control applications.

Optimal Control of Service in Tandem Queues

ZVI ROSBERG, PRAVIN P. VARAIYA, FELLOW, IEEE, AND JEAN C. WALRAND, MEMBER, IEEE

***Abstract*—Customers arrive in a Poisson stream into a network consisting of two $M/M/1$ service stations in tandem. The service rate $u \in [0, a]$ at station 1 is to be selected as a function of the state (x_1, x_2) where x_i is the number of customers at station i so as to minimize the expected total discounted or average cost corresponding to the instantaneous cost $c_1x_1 + c_2x_2$. The optimal policy is of the form $u = a$ or $u = 0$ according as $x_1 < S(x_2)$ or $x_1 \geqslant S(X_2)$ and S is a switching function. For the case of discounted cost, the optimal process can be nonergodic, but it is ergodic for the case of average cost.**

I. Introduction

The search for the optimal control policy in queueing control problems can be simplified if it is known that the optimal policy has a special structure, e.g., "bang-bang" or "switchover." Queueing models consisting of a single service station with controllable service times have been studied for various cost functions and shown to possess switchover optimal policies.

An extensive bibliography on queueing control models can be found in Grabill *et al.* [4]. The analysis of an $M/M/1$ model by Lippman [5], [6] extends earlier results of Grabill [3]. In his study of an $M/G/1$ model, Gallisch [2] gives conditions on the cost function and service time distribution which imply switchover optimal policies. He extends results obtained in Schassberger [8] where the service cost is linear and the waiting time cost is not allowed, and in Tijms [9] where the waiting time cost is linear. Formulas and properties of the long-run average cost under switchover policies are derived in Tijms *et al.* [10]. They unify several specific models for controlling service and arrival rates in an $M/G/1$ model. The formulas derived there for the cost may be used to obtain operating characteristics of the system. For an $M/G/1$ system, Mitchell [7] and Doshi [1] derive similar properties of ϵ-optimal control laws when the cost is convex and separable in service rate and queue length.

There does not seem to be any study of optimal policies when there are two or more connected service stations. This paper analyzes the simplest such case where customers in a Poisson stream enter a network consisting of two exponential servers in tandem. The service rate $u \in [0, a]$ at station 1 is to be selected as a function of the state (x_1, x_2) where x_i is the number of customers in station i. The instantaneous cost is linear in the waiting times at the two stations $c_1x_1 + c_2x_2$, and we show that the policies which minimize the total expected discounted or long-run average cost are switchover. That is, the optimal policy is characterized by a switching curve S: $u = a$ or $u = 0$ accordingly as $x_1 \geqslant S(x_2)$ or $x_1 < S(x_2)$. The optimal S can be interpreted as the condition that $x_1 \cong S(x_2)$ iff in the state (x_1, x_2), the marginal increase in the expected cost is the same whether a marginal customer is added to the queue at 1 or at 2. This interpretation also explains the fact that the optimal S is monotonically increasing.

The analysis of the problem differs in two respects from the previously mentioned studies of a single server. The first concerns the convexity of the value function. In a single server model, the state space is one dimensional, and showing convexity of the value function using the optimality equations is usually sufficient for proving switchover of

Manuscript received January 22, 1981; revised June 29, 1981. Paper recommended by S. I. Marcus, Past Chairman of the Stochastic Control Committee. This work was supported in part by the National Science Foundation under Grant EGN-7903879 A01.

Z. Rosberg is with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa, Israel.

P. P. Varaiya is with the Department of Electrical Engineering and Computer Science and the Electronics Research Laboratory, University of California, Berkeley, CA 94720.

J. C. Walrand is with the Department of Electrical Engineering and Computer Science and the Electronics Research Laboratory, University of California, Berkeley, CA 94720.

0018-9286/82/0600-0600$00.75 ©1982 IEEE

the optimal policy. In our case, the state space is two dimensional, and the proof of convexity does not directly follow from the optimality equations. Our proof is based upon constructing an equivalent linear programming problem and deducing convexity from the duality theory. The second difference concerns the analysis of the case of long-run average cost. In single server models, it is usually trivial to show ergodicity of the optimal process for the discounted case. The average cost is then treated as a limit of the discounted case as in Lippman [5]. In our problem, the optimal process for the discounted case can be non-ergodic (in fact, all states can be transient), and so the case of average cost cannot be treated as the limit of the discounted case. Instead, we follow the unusual procedure of moving to the average cost from the finite horizon problem.

The paper is organized as follows. In the next section, the continuous-time problem is transformed into one with discrete time by looking at an embedded Markov chain, and the optimality conditions are written down. In Section III, the equivalent linear programming is introduced to prove convexity of the value function. The optimality conditions and convexity are used in Section IV to show that the optimal policy is switchover for the finite horizon and discounted problems. The average cost case is treated in Section V where a separate argument is given to show ergodicity.

II. The Equivalent Discrete-Time Problem

Customers arrive in station 1 in a Poisson stream with constant rate λ. The rate at time t of the exponential server in station 1 can be selected to be any number u_t in $[0, a]$. Upon completing service at 1, a customer joins the queue at station 2 which is served by another exponential server with constant service rate μ. Let x_{it} be the number of customers at time t in station i, the customer in service included, and let $x_t = (x_{it}, x_{2t})$ denote the state at t. u_t is to be selected knowing x_t.

The cost incurred per unit time in state x is $c'x = c_1x_1 + c_2x_2$ where $c_1 > 0$, $c_2 > 0$ are fixed. Let $\alpha \geq 0$ be the interest rate used for discounting future cost, i.e., the present value of cost c incurred at time t is $ce^{-\alpha t}$. Let $V_t^\alpha(x)$ be the minimum achievable cost when the time horizon is $t \geq 0$ and $x_0 = x$.

Dynamic programming considerations lead to the following optimality conditions for V_t^α.

$$
\begin{aligned}
&V_{t+dt}^\alpha(x)\\
&= c'xdt + e^{-\alpha dt}\\
&\quad \cdot \inf_{0 \leq u \leq a} \{\lambda dt V_t^\alpha(Ax) + \mu dt V_t^\alpha(Dx) + udt V_t^\alpha(Tx)\\
&\quad + [1 - (\lambda + \mu + u)dt] V_t^\alpha(x) + o(dt)\},\\
&= c'xdt + e^{-\alpha dt}\{\lambda dt V_t^\alpha(Ax) + \mu dt V_t^\alpha(Dx)\\
&\quad + [1 - (\lambda + \mu)dt] v_t^\alpha(x)\}\\
&\quad + e^{-\alpha dt} \inf_{0 \leq u \leq a} udt[V_t^\alpha(Tx) - V_t^\alpha(x)] + o(dt). \quad (2.1)
\end{aligned}
$$

In (2.1), A, D, T are functions representing an arrival at station 1, a departure from station 2, and a service completion at 1. That is, $A(x_1, x_2) = (x_1 + 1, X_2)$; $D(x_1, x_2) = (x_1, (x_2 - 1)^+)$; $T(x_1, x_2) = (x_1 - 1, x_2 + 1)$ or (x_1, x_2) accordingly as $x_1 > 0$ or $x_1 = 0$.

Observe that by writing the optimality conditions in this way, we are adopting the convention that an idle server at stations 1 or 2 (idle by choice or because of lack of customers) is serving a dummy customer who incurs no waiting cost and who never leaves the station. This is permissible because of the memoryless nature of the exponential service time.

From (2.1) it follows that the policy given by $u_s = 0$ if $V_{t-s}^\alpha(Tx_s) > V_{t-s}^\alpha(x_s)$ and $u_s = a$ if $V_{t-s}^\alpha(Tx_s) \leq V_{t-s}^\alpha(x_s)$ is optimal. Hence, we can limit attention to policies which are "bang-bang," i.e., which take values in $\{0, a\}$. To convert the problem into one in discrete time, we henceforth limit ourselves further to those bang-bang policies which change values only at a transition epoch, including as transitions those due to service completions for dummy customers. Call this class of policies P. For the infinite horizon problem, this is an inessential restriction since the optimal policy is stationary, and hence it is in P. For the finite horizon, this is a restriction.

The transition epochs $0 = t_0 < t_1 < t_2 \cdots < t_n \cdots$ are the same for all policies in P. In fact, the interepoch intervals are independent and have the same distribution, $\text{Prob}\{t_{k+1} - t_k > t\} = \exp - t(\lambda + \mu + a)$. The cost incurred by a policy (in P) over the random interval $[0, t_n]$ and with initial state x is

$$
\begin{aligned}
E\int_0^{t_n} e^{-\alpha t} c' x_t dt &= E_x \sum_0^{n-1} \int_{t_k}^{t_{k+1}} e^{-\alpha t} c' x_{t_k} dt\\
&= E_x \Sigma \left\{ c' x_{t_k} E \int_{t_k}^{t_{k+1}} e^{-\alpha t} dt \right\}\\
&= E_x \Sigma c' x_{t_k} \frac{1}{\alpha} E[e^{-\alpha t_k} - e^{-\alpha t_{k+1}}]. \quad (2.2)
\end{aligned}
$$

Since the $t_{k+1} - t_k$'s are i.i.d., therefore $Ee^{-\alpha t_k} = \beta^k$ where

$$
\begin{aligned}
\beta = Ee^{-\alpha t_1} &= \int_0^\infty e^{-\alpha t}(\lambda + \mu + a) e^{-(\lambda + \mu + a)t} dt\\
&= \frac{\lambda + \mu + a}{\alpha + \lambda + \mu + a}. \quad (2.3)
\end{aligned}
$$

Substituting, we see that the cost (2.2) equals

$$
\frac{1 - \beta}{\alpha} E_x \sum_{k=0}^{n-1} \beta^k c' x_{t_k}, \quad (2.4)
$$

provided that $\beta < 1$, whereas if $\beta = 1$, then it equals $(\lambda + \mu + a)^{-1} \Sigma E_x c' x_k$. It is convenient to ignore the constant factor and take as the cost

$$
E_x \sum_{k=0}^{n-1} \beta^k c' x_{t_k}
$$

which is valid for $0 \leq \beta \leq 1$. Writing $x_k = x_{t_k}$, we see that the search for the optimal policy in P over the random interval $[0, t_n]$ is equivalent to finding the best policy over

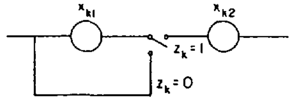

Fig. 1. The Markov decision problem.

the horizon $[0, n]$ for the following discrete-time Markov decision problem.

The state space is N^2 where $N=\{0,1,2,\cdots\}$, the action space is $\{0,1\}$, and the one-step transition probabilities are

$$p(x_{k+1}|x_k, z_{k+1})=\begin{cases}\lambda & \text{if } x_{k+1}=Ax_k\\ \mu & \text{if } x_{k+1}=Dx_k\\ a & \text{if } x_{k+1}=Tx_k,\ z_{k+1}=1\\ a & \text{if } x_{k+1}=x_k,\ z_{k+1}=0.\end{cases} \tag{2.5}$$

In (2.5), the (λ, μ, a) are proportional to the (λ, μ, a) in (2.1) and are normalized so that $\lambda+\mu+a=1$. $z_{k+1}=1$, respectively 0, corresponds to $u_t\equiv a$, respectively 0, during the interval $[t_k, t_{k+1}]$. $1-z_{k+1}$ can be interpreted as the probability of recycling a customer as in Fig. 1.

The cost is given by $E_x\Sigma^{n-1}\beta^k c'x_k$. Let V_n^β denote the minimum achievable cost. It is characterized by these optimality conditions: $V_1^\beta(x)=c'x$ and

$$\begin{aligned}V_{n+1}^\beta(x)&=c'x+\beta\min_{z\in\{0,1\}}\{\lambda V_n^\beta(Ax)+\mu V_n^\beta(Dx)\\&\quad+azV_n^\beta(Tx)+a(1-z)V_n^\beta(x)\}\\&=c'x+\beta\{\lambda V_n^\beta(Ax)+\mu V_n^\beta(Dx)\\&\quad+a\min(V_n^\beta(x),V_n^\beta(Tx))\}.\end{aligned} \tag{2.6}$$

Furthermore, when there are k steps to go and the current state is x, the optimal action is $z=0$ if $V_k^\beta(Tx)>V_k^\beta(x)$ and $z=1$ if $V_k^\beta(Tx)\leqslant V_k^\beta(x)$.

Clearly, $c'x\geqslant 0$, therefore $V_n^\beta(x)\leqslant V_{n-1}^\beta(x)$, and so this limit exists:

$$V_\infty^\beta(x)=\lim_{n\to\infty}V_n^\beta(x). \tag{2.7}$$

Lemma 2.1: $V_\infty^\beta<\infty$ for $\beta<1$.

Proof: Let $|x|=|x_1|+|x_2|$ and $|c|=\max(c_1,c_2)$. For any initial state x and any policy, the state x_k at time k must satisfy $|x_k|\leqslant|x|+k$. Hence,

$$\begin{aligned}V_n^\beta(x)&\leqslant\sum_{k=0}^{n-1}\beta^k|c|(|x|+k)\\&<\frac{|c|\,|x|}{1-\beta}+\frac{|c|\beta}{(1-\beta)^2}<\infty.\end{aligned} \tag{2.8}$$

□

It follows from Theorem 1 in Lippman [5] that for $\beta<1$, the minimum cost for the infinite horizon

$$V^\beta(x)=\min\sum_{k=0}^{\infty}E_x c'x_k$$

is achieved by a stationary policy. Moreover, V^β is the unique solution to the optimality condition

$$\begin{aligned}V^\beta(x)=c'x+\beta\{\lambda V^\beta(Ax)+\mu V^\beta(Dx)\\+a\min(V^\beta(x),V^\beta(Tx))\}.\end{aligned} \tag{2.9}$$

From (2.6), (2.7), and the uniqueness of V^β, it follows that

$$V^\beta=V_\infty^\beta. \tag{2.10}$$

III. $V_n^\beta(x)$ is Convex

We have been unable to show that $V_n^\beta(x)$ is convex by using the relation (2.6). Instead, we give a linear programming problem equivalent to the Markov decision problem and then deduce convexity by duality. This technique appears novel and may be useful in studying other queueing problems.

The basic sample space for the Markov decision problem is denoted Ω^n, and it consists of all sequences $\omega^n=(\omega_1,\cdots,\omega_n)$ where $\omega_k\in\{A,T,D\}$; $\omega_k=A$, respectively T or D, accordingly as the transition at step k is an arrival, respectively a service completion at stations 1 or 2, including service completion for dummy customers. Let p denote the probability distribution on Ω^n. Under p, the ω_k's are i.i.d. and $\omega_k=A$, T, or D with probability λ, a, or μ. Let $\underline{F}_k$ be the field on Ω^n generated by $\omega^k=(\omega_1,\cdots,\omega_k)$. A *process* is any sequence of random variables $f=(f_1,\cdots,f_n)$ such that f_k is adapted to $\underline{F}_k$. Hence, we can and will regard f_k as a function on Ω^k. Let

$$\xi(\omega)=\begin{cases}(1,0) & \text{if } \omega=A\\ (-1,1) & \text{if } \omega=T\\ (0,-1) & \text{if } \omega=D\end{cases}$$

and let $\xi=(\xi_1,\cdots,\xi_n)$ be the process given by $\xi_k(\omega^k)=\xi(\omega_k)$.

A *policy* is any process $z=(z_1,\cdots,z_n)$ satisfying

$$z_k(\omega^k)\begin{cases}=1 & \text{if } \omega_k=A\\ \in[0,1] & \text{if } \omega_k=T\\ \in[0,1] & \text{if } \omega_k=D.\end{cases} \tag{3.1}$$

Let Z be the set of all policies and Z_I the subset of policies which are integer valued.

The *trajectory* corresponding to a policy z (and initial state x) is the process $\hat{x}=(x_1,\cdots,x_n)$ with values in R^2, defined recursively by

$$\begin{aligned}x_0&=x,\\x_k(\omega^k)&=x_{k-1}(\omega^{k-1})+z_k(\omega^k)\xi_k(\omega^k),\qquad k>0.\end{aligned} \tag{3.2}$$

The trajectory is said to *nonnegative*, and we write $z\geqslant 0$ if for all ω^k, $x_k(\omega^k)\geqslant 0$.

The cost incurred by policy z and initial state x is defined to be

$$J_{n+1}(z,x)=E_x\sum_{k=0}^{n}\beta^k c'x_k.$$

From (3.2),

$$x_k(\omega^k)=x+\sum_{j=1}^{k}z_j(\omega^j)\xi_j(\omega^j),\qquad k>0. \tag{3.3}$$

Using this, we obtain after some manipulations

$$J_{n+1}(z,x)=(1+\beta+\cdots+\beta^n)c'x+\sum_{k=1}^{n}\sum_{\omega^k\in\Omega^k}\gamma_k(\omega^k)z_k(\omega^k)$$

where $(\gamma_1,\cdots,\gamma_n)$ is the process

$$\gamma_k(\omega^k)=\sum_{j=k}^{n}\left\{\sum_{\omega_{k+1},\cdots,\omega_j}p\left(\omega^k,\omega_{k+1},\cdots,\omega_j\right)\right\}\beta^j c'\xi(\omega_k). \tag{3.4}$$

When x and z are integer valued, then so is the trajectory $\hat{x}$; if, moreover, $\hat{x}\geqslant 0$, then it does describe the evolution of the two tandem queues under the policy z and $J_n(z,x)$ is the n-step cost. (If x and u are not integer valued, but if $x\geqslant 0$, then we can interpret $\hat{x}$ as the evolution of the queueing system in which there are "fractional" customers.) The value function for the Markov decision problem is also given by

$$V_n^\beta(x)=\min\{J_n(z,x)\,|\,z\in Z_I,\ \hat{x}\geqslant 0\}. \tag{3.5}$$

Define

$$W_n^\beta(x)=\min\{J_n(z,x)\,|\,z\in Z,\ \hat{x}\geqslant 0\}. \tag{3.6}$$

Theorem 3.1 is an immediate consequence of Lemmas 3.1 and 3.2.

Theorem 3.1: $V_n^\beta(x)$ is convex for $x\in N^2$, i.e., $x\geqslant 0$ and integer valued.

Lemma 3.1: $W_n^\beta(x)$ is a convex function for $x\geqslant 0$.

Lemma 3.2: $V_n^\beta(x)=W_n^\beta(x)$ for $x\in N^2$.

Proof of Lemma 3.1: For $x\geqslant 0$, let $M(\hat{x})=W_{n+1}^\beta(x)-(n+1)c'x$. From (3.1), (3.3), (3.4), and (3.6), we get

$$M(x)=\min\sum_{k=1}^{n}\sum_{\omega^k}\gamma_k(\omega^k)z_k(\omega^k)$$

$$\text{s.t.}\quad z_k(\omega^k)=1\qquad \text{if } \omega_k=A, \tag{3.7}$$

$$0\leqslant z_k(\omega^k)\leqslant 1\qquad \text{if } \omega_k=T \text{ or } D, \tag{3.8}$$

$$x_k(\omega^k,z):=x+\sum_{j=1}^{k}z_j(\omega^j)\xi_j(\omega^j)\geqslant 0,\qquad \omega^k\in\Omega^k,\ k\geqslant 1. \tag{3.9}$$

This is a linear programming (LP) problem in the finite array of variables $z=\{z_k(\omega^k)\,|\,\omega^k\in\Omega^k,\ 1\leqslant k\leqslant n\}$. Since x enters linearly in the constraint equation, therefore $M(x)$ is convex (see, e.g., [11, p. 93, Lemma 3], and Lemma 3.1 is proved.

Proof of Lemma 3.2: By the duality theory [11, p. 98, Theorem 2], z^* is an optimal solution of the LP above if and only if there exist dual variables $\lambda_k^*(\omega^k)\in R^2$, $\lambda_k^*(\omega^k)\geqslant 0$, $\omega^k\in\Omega^k$, $k>1$ [and corresponding to the nonnegativity constraints (3.9)] such that conditions 1), 2), and 3) below hold.

1) z^* is an optimal solution of this LP:

$$\min\sum_{k=1}^{n}\sum_{\omega^k}\gamma_k(\omega^k)z_k(\omega^k)-\sum_{k=1}^{n}\lambda^*(\omega^k)'\left[x+\sum_{j=1}^{k}z_j(\omega^j)\xi_j(\omega^j)\right]$$

s.t. (3.7) and (3.8).

2) (Feasibility): $x_k(\omega^k,z^*)\geqslant 0$, $\omega^k\in\Omega^k$, $k>1$.

3) (Complementary slackness): If $\lambda_{ki}^*(\omega^k)>0$, then $x_{ki}(\omega^k,z^*)=0$, $i=1,2$, $k>1$.

The cost function in 1) may be rewritten as

$$\sum_{k=1}^{n}\sum_{\omega^k}\left[\gamma_k(\omega^k)-\left(\sum_{j=k}^{n}\lambda_j^*(\omega^j)'\xi_j(\omega^j)\right)\right]z_k(\omega^k)+C$$

where C is a constant independent of z. Hence, condition 1) can be more conveniently rewritten as 1′).

1′) If $\omega_k=A$, then $z^*(\omega^k)=1$; if $\omega_k=T$ or D, then

$$z^*(\omega^k)\begin{cases}=1 & \text{if } \alpha_k:=\gamma_k(\omega^k)-\sum_{j=k}^{n}\lambda_j^*(\omega^j)'\xi_j(\omega^j)<0\\ =0 & \text{if } \alpha_k>0\\ \in[0,1] & \text{if } \alpha_k=0.\end{cases}$$

Henceforth, suppose that $x\in N^2$. Let z^*,λ^* satisfy 1′), 2), and 3). We will use z^* to construct an integer-valued policy z which also satisfies these conditions.

Proposition 1: Let $X=\{z_1\xi(T)+z_2\xi(D)\,|\,-(1/2)<z_i\leqslant(1/2),\ i=1,2\}$. Then

$$X+\{z\xi(\omega)\,|\,0\leqslant z\leqslant 1\}\subset X\cup\{X+\xi(\omega)\},\qquad \omega=T \text{ or } D. \tag{3.10}$$

Proof: The proof is by simple verification. □

Proposition 2: There is an integer-valued policy $z=\{z_k(\omega^k)\}$ such that $z_k(\omega^k)=z_k^*(\omega^k)$ whenever the latter is integer valued, and

$$\Delta_k:=\left[x_k(\omega^k,z^*)-x_k(\omega^k,z)\right]\in X\qquad \text{for all } \omega^k,\ k>1.$$

Proof: Suppose that for some $k\geqslant 0$, the $z_j(\omega^j)$'s, $j\geqslant k$ have been selected and $\Delta_k\in X$. Now

$$\Delta_{k+1}=\Delta_k+z_{k+1}^*(\omega^{k+1})\xi(\omega_{k+1})-z_{k+1}(\omega^{k+1})\xi(\omega_{k+1}).$$

If $z_{k+1}^*(\omega^{k+1})$ is integer valued, then take $z_{k+1}(\omega^{k+1})=z_{k+1}^*(\omega^{k+1})$, and then certainly $\Delta_{k+1}\in X$. Otherwise, by (3.10), either $[\Delta_k+z^*(\omega^{k+1})\xi(\omega_{k+1})]$ is in X and then take $z_{k+1}(\omega^{k+1})=0$ or $[\Delta_k+z^*(\omega^{k+1})\xi(\omega_{k+1})]\in X+\xi(\omega_{k+1})$ and then take $z_{k+1}(\omega^{k+1})=1$. In either case, $\Delta_{k+1}\in X$, and the proposition follows by induction. □

Lemma 3.2 follows from the next proposition.

Proposition 3: The integer-valued policy z constructed above satisfies 1′), 2), and 3).

Proof: Since $z_k(\omega^k)=z_k^*(\omega^k)$ whenever the latter is integer valued, condition 1′) is satisfied.

To prove 2), it must be shown that $y:=x_k(\omega^k,z)\geqslant 0$. Suppose, to the contrary, that $y\ngeqslant 0$. Since y is integer valued, therefore for at least one $i=1$ or 2, $y_i\leqslant -1$. Let $y^*:=x_k(\omega^k,z^*)$. By Proposition 2, and recalling the definition of $\xi(T)$ and $\xi(D)$,

$$y^*\in y+\left\{z_1(-1,1)+z_2(0,-1)\,|\,-\frac{1}{2}<z_i\leqslant\frac{1}{2}\right\}$$
$$=\left\{(y_1-z_1,y_2+z_1-z_2)\,|\,-\frac{1}{2}<z_i\leqslant\frac{1}{2}\right\}.$$

Now if $y_1\leqslant -1$, then $y_1-z_1<0$ if $-(1/2)<z_1$ and so $y_1^*<0$; and if $y_2\leqslant -1$, then $y_2+z_1-z_2<0$ if $z_1\leqslant(1/2)$ and $-(1/2)<z_2$ and so $y_2^*<0$. In either case, if $y\ngeqslant 0$, then $y^*\ngeqslant 0$ which contradicts the hypothesis that z^* satisfies 2). Thus, z satisfies 2) as well.

Finally, to prove 3), it is evidently sufficient to show that $y_i=0$ whenever $y_i^*=0$ where these are defined as above. By Proposition 2 again,

$$y\in\left\{(y_1^*+z_1,y_2^*-z_1+z_2)\,|\,-\frac{1}{2}<z_i\leqslant\frac{1}{2}\right\}.$$

If $y_1^*=0$, then $y_1\in(-(1/2),(1/2)]$, and so being integer valued, $y_1=0$; and if $y_2^*=0$, then $y_2\in(-1,1)$, and so being integer valued, $y_2=0$. □

Corollary 3.1: For $\beta<1$, $V^\beta=V_\infty^\beta$ is convex.

Proof: The proof is immediate from Theorem 3.1 and (2.7). □

IV. The Optimal Switching Function

In this section, conditions (2.6), (2.9), and the convexity of V_n^β are used to show that the optimal policy is switchover. For the infinite horizon problem, only the discounted case $\beta<1$ is considered here.

Fix $\beta\leqslant 1$. Let $V_n=V_n^\beta$. From (2.6),

$$V_1(x)=c'x,\tag{4.1}$$
$$V_2(x)=(1+\beta)c'x+\beta\lambda c_1-\beta\mu c_2 1(x_2>0)+\beta a(c_2-c_1)1(c_2\leqslant c_1,x_1\geqslant 0)\tag{4.2}$$

where $1(\cdot)$ denotes the indicator function. To evaluate V_n for $n\geqslant 2$, define

$$U_n(x)=V_n(Tx)-V_n(x).\tag{4.3}$$

Observe, with the aid of (2.6), that when there are n steps to go, the optimal action is

$$z=\begin{cases}0 & \text{if } U_n(x)>0\\ 1 & \text{if } U_n(x)\leqslant 0.\end{cases}\tag{4.4}$$

Also from (2.6),

$$U_{n+1}(x)=(c_2-c_1)+\beta\lambda U_n(Ax)+\beta\mu[V_n(DTx)-V_n(Dx)]+\beta a\phi_n(x)\tag{4.5}$$

where

$$\phi_n(x)=V_n(Tx)\wedge V_n(T^2x)-V_n(x)\wedge V_n(Tx)\tag{4.6}$$

and $f\wedge g=\min(f,g)$ for any two numbers f and g.

It is clear, and can be proved easily by an argument based on stochastic dominance, that the total cost is an increasing function of the initial state.

Lemma 4.1: $V_n^\beta(x_1,x_2)$ is nondecreasing in x_i, $i=1,2$.

If $c_1\geqslant c_2$, that is, the waiting cost at station 2 is not greater than at station 1, then evidently the optimum policy is $z_k\equiv 1$.

Lemma 4.2: If $c_1\geqslant c_2$, then $z_k\equiv 1$ is an optimum policy.

Henceforth, it is assumed that $c_2>c_1$.

Lemma 4.3: $U_n(x_1,x_2)$ is increasing in x_2 and decreasing in x_1 for $x_1>0$, $x_2\geqslant 0$ (increasing means nondecreasing, similarly for decreasing).

Proof: From (4.1), (4.2), and (4.3),

$$U_1(x)=c_2-c_1\tag{4.7}$$
$$U_2(x)=(1+\beta)(c_2-c_1)-\beta\mu c_2 1(x_2=0)\tag{4.8}$$

so the assertion is true for $n=1,2$. Suppose that it is true for n. The first two terms in the formula (4.5) clearly have the indicated property. The third term is

$$\beta\mu[V_n(DTx)-V_n(Dx)]=\begin{cases}U_n(Dx) & \text{if } x_1>0,\ x_2>0\\ V_n(x_1-1,0)-V_n(x_1,0) & \text{if } x_1>0,\ x_2=0.\end{cases}\tag{4.9}$$

By the induction hypothesis, $U_n(Dx)$ has the indicated property for $x_1>0$, $x_2>0$. By Theorem 3.1 and Lemma 4.1, $V_n(x_1,0)$ is an increasing convex function of x_1, and so $V_n(x_1-1,0)-V_n(x_1,0)$ is decreasing in x_1 for $x_1>0$. It only remains to check that (4.9) is larger for $(x_1,1)$ than for $(x_1,0)$, i.e., that

$$U_n(D(x_1,1))=V_n(x_1-1,1)-V_n(x_1,0)\geqslant V_n(x_1-1,0)-V_n(x_1,0).$$

But this inequality follows from Lemma 4.1. The last term in (4.5) can be rewritten as

$$\beta a\{[V_n(T^2x)-V_n(Tx)]\wedge 0+[V_n(Tx)-V_n(x)]\vee 0\}=\beta a[U_n(Tx)\wedge 0+U_n(x)\vee 0],$$

and by the hypothesis, this expression also has the indicated property. The lemma follows by induction. □

Define the *switching function* $S_n(x_1)$ by

$$S_n(x_1)=\min\{x_2\geqslant 0\,|\,U_n(x_1,x_2)>0\}$$
$$=\infty \text{ if } U_n(x_1,x_2)\leqslant 0 \quad \text{for all } x_2.$$

It will be shown in Lemma 4.5 that S_n is always finite.

Corollary 4.1: $S_n(x_1)$ is increasing in x_1 for $x_1 \geq 1$. When there are n steps to go, the optimal action is $z=0$ or 1 accordingly as $x_2 \geq S_n(x_1)$ or $x_2 < S_n(x_1)$.

Proof: The first assertion follows from the fact that $U_n(x_1, x_2)$ is increasing in x_2, and the second follows from (4.4) and the fact that $U_n(x_1, x_2)$ is decreasing in x_1. □

Next we study the behavior of S_n as n increases.

Lemma 4.4: For $n \geq 1$,

$$U_n(x) \leq 0 \Rightarrow U_{n+1}(x) \leq U_n(x), \quad (4.10)$$

$$V_n(Ax) - V_n(x) \leq V_{n+1}(Ax) - V_{n+1}(x), \quad (4.11)$$

$$V_n(x) - V_n(Dx) \leq V_{n+1}(x) - V_{n+1}(Dx). \quad (4.12)$$

Proof: Consider $n=1$. If $x_1=0$, then $U_1(x)=U_2(x)=0$, and (4.10) holds. If $x_1>0$, then by (4.7), $U_1(x)=c_2-c_1>0$, so (4.10) holds trivially. By (4.1) and (4.2),

$$V_1(Ax)-V_1(x)=c_1 \leq (1+\beta)c_1 = V_2(Ax)-V_2(x),$$
$$V_1(x)-V_1(Dx)=c_2 1(x_2>0) \leq (1+\beta)c_2 1(x_2>0) - \beta\mu c_2 1(x_2=1) = V_2(x)-V(Dx),$$

and so the assertion is true for $n=1$. Suppose that the assertion is true for $n-1$. We will prove it for n.

Consider (4.10), and suppose that $U_n(x) \leq 0$. By (4.5),

$$\begin{aligned} U_n(x)-U_{n+1}(x) &= \beta\lambda[U_{n-1}(Ax)-U_n(Ax)] \\ &+ \beta\mu\{[V_{n-1}(DTx)-V_{n-1}(Dx)] \\ &-[V_n(DTx)-V_n(Dx)]\} \\ &+\beta a[\phi_{n-1}(x)-\phi_n(x)]. \end{aligned} \quad (4.13)$$

Now $[U_{n-1}(Ax)-U_n(Ax)] \geq 0$. To see this, note that $U_n(x) \leq 0$ implies that $U_n(Ax) \leq 0$ since U_n is decreasing in x_1. Therefore, if $U_{n-1}(Ax)>0$, then certainly $U_{n-1}(Ax) - U_n(Ax) \geq 0$, whereas if $U_{n-1}(Ax) \leq 0$, then the same conclusion follows from the induction hypothesis. The coefficient of $\beta\mu$ in (4.13) is also nonnegative. Indeed, if $x_2>0$, then this coefficient is $[U_{n-1}(Dx)-U_n(Dx)]$ and the same argument applies, whereas if $x_2=0$, this coefficient is $[V_{n-1}(x_1-1,0)-V_{n-1}(x_1,0)]-[V_n(x_1-1,0)-V_n(x_1,0)]$, and this is nonnegative by the induction hypothesis. The coefficient of βa in (4.13) can be rewritten as

$$\begin{aligned} &[U_{n-1}(Tx)\wedge 0 + U_{n-1}(x)\vee 0] - [U_n(Tx)\wedge 0 + U_n(x)\vee 0] \\ &\quad = U_{n-1}(Tx)\wedge 0 + U_{n-1}(x)\vee 0 - U_n(Tx)\wedge 0 \end{aligned} \quad (4.14)$$

since $U_n(x) \leq 0$. The induction hypothesis implies that $U_{n-1}(Tx)\wedge 0 \geq U_n(Tx)\wedge 0$, and so the coefficient of βa is nonnegative. Thus, (4.10) is true for n.

Next consider (4.11). Using (2.6),

$$\begin{aligned} &V_{n+1}(Ax)-V_{n+1}(x) \\ &\quad = c_1 + \beta\lambda[V_n(A^2x)-V_n(Ax)] + \beta\mu[V_n(DAx)-V_n(Dx)] \\ &\qquad + \beta a[V_n(Ax)\wedge V_n(TAx) - V_n(x)\wedge V_n(Tx)]. \end{aligned}$$

A similar expression is valid for $V_n(Ax)-V_n(x)$. By the induction hypothesis,

$$V_n(A^2x)-V_n(Ax) \geq V_{n-1}(A^2x)-V_{n-1}(Ax),$$
$$V_n(DA_x)-V_n(Dx) \geq V_{n-1}(DAx)-V_{n-1}(Dx).$$

Therefore, to show (4.11), it only remains to show that

$$v_n(x) \geq v_{n-1}(x) \quad (4.15)$$

where $v_n(x)=V_n(Ax)\wedge V_n(TAx) - V_n(x)\wedge V_n(Tx)$. Since for real numbers a, b, c, d we have $a\wedge b - c\wedge d = a - d - (a-b)\vee 0 - (c-d)\wedge 0$, we may write

$$v_n(x) = V_n(TAx)-V_n(x)-U_n(Ax)\vee 0 - U_n(x)\wedge 0.$$

Since $x = DTAx$, the induction hypothesis yields

$$V_n(TAx)-V_n(x) \geq V_{n-1}(TAx)-V_n(x), \quad (4.16)$$

$$U_n(x)\wedge 0 \leq U_{n-1}(x)\wedge 0. \quad (4.17)$$

Now, if $U_{n-1}(Ax) \leq 0$, then by the induction hypothesis, $U_n(Ax) \leq 0$ also, which together with (4.16) and (4.17) gives (4.15); if $U_{n-1}(Ax)>0$ and $U_n(Ax) \leq 0$, then again we get (4.15); finally, if $U_{n-1}(Ax)>0$ and $U_n(Ax)>0$, then $U_{n-1}(x)>0$ and $U_n(x)>0$ (since U_n and U_{n-1} are decreasing in x_1), and so again

$$\begin{aligned} v_n(x) &= V_n(TAx)-V_n(x)-U_n(Ax) = V_n(Ax)-V_n(x) \\ &\geq V_{n-1}(Ax)-V_{n-1}(x) = v_{n-1}(x). \end{aligned}$$

Finally, consider (4.12). Using (2.6), it can be seen that (4.12) holds if

$$w_n(x) \geq w_{n-1}(x) \quad (4.18)$$

where $w_n(x)=V_n(x)\wedge V_n(Tx)-V_n(Dx)\wedge V_n(TDx)$. Now

$$\begin{aligned} w_n(x) &= V_n(x)-V_n(Dx)+[V_n(Tx)-V_n(x)]\wedge 0 \\ &\quad -[V_n(TDx)-V_n(Dx)]\wedge 0. \end{aligned}$$

By the induction hypothesis,

$$V_n(x)-V_n(Dx) \geq V_{n-1}(x)-V_{n-1}(Dx),$$
$$[V_n(TDx)-V_n(Dx)]\wedge 0 \leq [V_{n-1}(TDx)-V_{n-1}(Dx)]\wedge 0.$$

Hence, (4.18) certainly holds if $U_{n-1}(x) \leq 0$ or $U_n(x) \geq 0$, whereas if $U_{n-1}(x)>0$ and $U_n(x)<0$, then $U_n(Dx)<0$ (since U_n is increasing in x_2), and so

$$w_n(x) = V_n(Tx)-V_n(TDx),$$
$$w_{n-1}(x) = V_{n-1}(x)-V_{n-1}(Dx)\wedge V_{n-1}(TDx).$$

Now if $U_{n-1}(Dx)<0$, then

$$\begin{aligned} w_{n-1}(x) &= V_{n-1}(x)-V_{n-1}(TDx) \\ &= V_{n-1}(Tx)-V_{n-1}(TDx)-U_{n-1}(x) \\ &\leq V_{n-1}(Tx)-V_{n-1}(TDx) \qquad \text{since } U_{n-1}(x)>0, \\ &\leq V_n(Tx)-V_n(TDx) \qquad \text{by the induction hypothesis}, \\ &= w_n(x), \end{aligned}$$

and similarly, if $U_{n-1}(Dx)>0$, then

$$w_{n-1}(x)=V_{n-1}(x)-V_{n-1}(Dx)$$
$$\leqslant V_n(x)-V_n(Dx) \qquad \text{by the induction hypothesis,}$$
$$\leqslant V_n(Tx)-V_n(TDx)=w_n(x).$$

The last inequality comes from the fact that $U_n(Dx)\leqslant U_n(x)$ since U_n is increasing in x_2.

The lemma now follows by induction. □

Corollary 4.2: The switching curve $S_n(x_1)$ is increasing in n.

Proof: If $x_2\leqslant S_n(x_1)$, then $U_n(x_1,x_2)\leqslant 0$, and so $U_{n+1}(x_1,x_2)\leqslant 0$, which implies that $x_2\leqslant S_{n+1}(x_1)$. Hence, $S_n(x_1)\leqslant S_{n+1}(x_1)$. □

Lemma 4.5: $S_n(x_1)\leqslant(n-2)\vee 0$ for all x_1.

Proof: Because of Corollary 4.1, we may limit ourselves to $x_1\geqslant 1$.

By Lemma 4.3, it is enough to show that $U_n(x_1,(n-2)\vee 0)>0$. For $n=1,2$, this follows from (4.7) and (4.8). Suppose that the assertion holds for some $n\geqslant 2$. Using (4.5),

$$\begin{aligned}U_{n+1}(x_1,n-1)&=c_2-c_1+\beta\lambda U_n(x_1+1,n-1)\\&\quad+\beta\mu[V_n(x_1-1,n-1)-V_n(x_1,n-2)]\\&\quad+\beta a\phi_n(x_1,n-1)\\&=c_2-c_1+\beta\lambda U_n(x_1+1,n-1)\\&\quad+\beta\lambda U_n(x_1,n-2)+\beta a\phi_n(x_1,n-1)\end{aligned}$$

and so it is enough to prove that $\phi_n(x_1,n-1)\geqslant 0$. By (4.6),

$$\begin{aligned}\phi_n(x_1,n-1)&=[V_n(T^2x)-V_n(Tx)]\wedge 0+U_n(x)\vee 0\\&\geqslant[V_n(T^2x)-V_n(Tx)]\wedge 0+U_n(x)\vee 0\\&\qquad\qquad\text{with } x=(x_1,n-1)\\&=U_n(x_1-1,n)\end{aligned}$$

which is positive by the induction hypothesis if $x_1-1\geqslant 1$ and also if $x_1=1$ because $U_n(0,n)=0$. □

Since we wish to consider different discount factors, we reintroduce the superscript β and use the notation $V_n^\beta, U_n^\beta, S_n^\beta$. Because of Corollary 4.2, the following limit exists:

$$S^\beta(x_1)=S_\infty^\beta(x_1)=\lim_{n\to\infty}S_n^\beta(x_1),$$

and by Corollary 4.2, it is increasing in x_1.

Theorem 4.1: An optimal policy for the infinite horizon discounted case $\beta<1$ is given by the stationary switchover policy $z(x)=0$ or 1 accordingly as $x_2\geqslant S^\beta(x_1)$ or $x_2<S^\beta(x_1)$.

Proof: Since $V_n^\beta\to V^\beta$, therefore $U^\beta(x)=\lim U_n^\beta(x)$. By (2.9), it is enough to show that

$$x_2\geqslant S^\beta(x_1)\Rightarrow U^\beta(x)\geqslant 0,$$
$$x_2<S^\beta(x_1)\Rightarrow U^\beta(x)\leqslant 0.$$

Now if $x_2\geqslant S^\beta(x_1)$, then $x_2\geqslant S_n^\beta(x_1)$ for all n, and so $U_n^\beta(x)\geqslant 0$; hence, $U^\beta(x)\geqslant 0$. Suppose that $x_2<S^\beta(x_1)$. Since the switching curves are integer valued and $S^\beta(x_1)=\lim S_n^\beta(x_1)$, therefore $x_2<S_n^\beta(x_1)$ for large n; hence, $U_n^\beta(x)\leqslant 0$ for large n, and so $U^\beta(x)\leqslant 0$. □

A surprising feature of the discounted case is that it is optimal to provide no service at station 1 when the queue length at station 2 exceeds a threshold.

Theorem 4.2: If $\beta<1$, then

$$S^\beta(x_1)\leqslant\min\left\{x_2\,|\,\beta^{x_2}\leqslant\frac{c_2-c_1}{c_2}\right\}.$$

Proof: Fix x with $x_1>0$. We claim that

$$\begin{aligned}U_n^\beta(x)&=V_n^\beta(Tx)-V_n^\beta(x)\\&\geqslant(c_2-c_1)\frac{1-\beta^{x_2\wedge n}}{1-\beta}-c_1\frac{\beta^{x_2\wedge n}-\beta^n}{1-\beta}\\&=v_n(x_2),\quad\text{say.}\end{aligned}\tag{4.19}$$

To see this, and using the notation of Section III, along a sample path ω^n, let $x_k(\omega^k)$ be the trajectory corresponding to the optimal policy and initial state $x_0=Tx$, and let $z_k(\omega^k)$ be the optimal control policy. Let $y_k(\omega^k)$ be the trajectory along the same sample path and corresponding to the same policy, but with initial condition $y_0=x$. Then

$$V_n^\beta(Tx)=E\sum_0^{n-1}\beta^kc'x_k\ \text{ and }\ V_n^\beta(x)\leqslant E\sum_0^{n-1}\beta^kc'y_k,$$

and so to prove (4.19), it is enough to show that

$$E\sum_0^{n-1}\beta^kc'x_k-E\sum_0^{n-1}\beta^kc'y_k\geqslant v_n(x_2).\tag{4.20}$$

To do this, define the stopping time

$$K=K(\omega^n)=\min\{k\geqslant 0\,|\,x_{k2}(\omega^k)=0\}\wedge n.$$

Then $x_{k2}>0$ for $k\leqslant K-1$, and a comparison of the two trajectories gives

$$\sum_0^{n-1}\beta^kc'x_k-\sum_0^{n-1}\beta^kc'y_k\geqslant(c_2-c_1)\sum_0^{K-1}\beta^k-c_1\sum_K^{n-1}\beta^k.$$

Since x_{k2} can decrease by at most one per step, therefore $K\geqslant x_2\wedge n$, and so

$$\sum_0^{n-1}\beta^kc'x_k-\sum_0^{n-1}\beta^kc'y_k\geqslant(c_2-c_1)\sum_0^{(x_2\wedge n)-1}\beta^k-c_1\sum_{x_2\wedge n}^{n-1}\beta^k$$

from which follows (4.20) by taking expectations. Hence, (4.19) is true. But then if $x_2\geqslant 0$ is such that $\beta^{x_2}\leqslant(c_2-c_1/c_2)$, then $U_n^\beta(x)>0$, and so $x_2\geqslant S_n^\beta(x_1)$. □

We have seen that if $\beta<1$, then the switching curves S_n^β increase to S^β and are all uniformly bounded. (See Fig. 2.) Let x_k^β, $k=0,1,\cdots$ be the Markov process corresponding to the optimal switching curve S^β. Since the service rate at station 1 is nonzero only when $x_2<S^\beta(x_1)$, therefore all states in $\{x\,|\,x_2\geqslant S^\beta(x_1)+1\}$ are transient. Observe from Theorem 4.2 that the service at station 1 stops whenever $x_{k2}^\beta\geqslant s$ where $s<\infty$ is independent of λ,μ. Hence, the

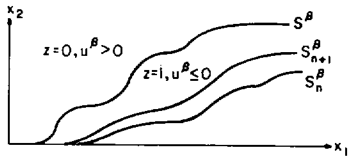

Fig. 2. Switching curves, $\beta<1$.

average effective service rate at station 1 is at most $a[1-b(\mu)]$ where $b(\mu)$ is the probability that x_2 exceeds s. If the process is ergodic, then one must have $a[1-b(\mu)]\geqslant\lambda$. However, $b(\mu)$ increases to 1 as μ decrease to λ. Therefore, there always exists μ sufficiently small, but with $\mu>\lambda$, such that the queue length in station $1x^\beta_{k_1}\to\infty$ with probability one. Thus, in the discounted case, the optimal process may be nonergodic; in fact, all states may be transient! The surprising result also suggests that the case of long-run average cost, whose study requires ergodicity, cannot be approached by the standard technique of taking limits as β increases to 1.

V. Long-Run Average Cost

Throughout this section, we consider the undiscounted case $\beta=1$, and so we write V_n, U_n, S_n instead of V_n^1, U_n^1, S_n^1. The next result contrasts sharply with Theorem 4.2.

Lemma 5.1: For every x_2, $S_n(x_1)>x_2$ for all (x_1,n) sufficiently large.

Proof: Suppose, in contradiction, that $S_n(x_1)<b$ for all x_1,n. We use the notation of Section III. Fix the horizon n and an initial state x. Let $z_k(\omega^k)$ be the optimal control policy and let $x_k(\omega^k)$ be the corresponding trajectory. Also, let $y_k(\omega^k)$ be the trajectory corresponding to the same policy, but with initial state Tx. Then

$$E\sum_0^{n-1}c'x_k=V_n(x),\quad E\sum_0^{n-1}c'y_k\geqslant V_n(Tx). \tag{5.1}$$

We wish to compare the two trajectories $\{x_k\}$ and $\{y_k\}$. Let K_i be the first time that queue i is empty:

$$K_i(\omega^n)=\min\{k\,|\,x_{k_i}(\omega^k)=0\}\wedge n.$$

If $K_1\leqslant K_2$, one can verify that

$$y_k=\begin{cases}Tx_k, & k=0,\cdots,K_1-1\\ x_k, & k\geqslant K_1,\end{cases}$$

and therefore,

$$\sum_0^{n-1}c'y_k-\sum_0^{n-1}c'x_k=(c_2-c_1)K_1,$$

whereas if $K_1>K_2$, then

$$y_k=\begin{cases}Tx_k, & k=0,\cdots,K_2-1\\ (x_{k1}-1,x_{k2}), & k=K_2,\cdots,K_1-1\\ (x_{k1},x_{k2}-1)\text{ or }x_k, & k\geqslant K_1,\end{cases}$$

and therefore,

$$\sum_0^{n-1}c'y_k-\sum_0^{n-1}c'x_k\leqslant(c_2-c_1)K_2-c_1(K_1-K_2).$$

Combining these two estimates gives

$$E\sum_0^{n-1}c'y_k-E\sum_0^{n-1}c'x_k\leqslant c_2E(K_1\wedge K_2)-c_1EK_1=\delta,\text{ say.} \tag{5.2}$$

We want to show that $\delta<0$ for some (x_1,n). It will be first shown that $E(K_2)<\overline{K}_2$ for some constant $\overline{K}_2$ and for all initial states $x=(x_1,b)$ and n. To see this, observe that, by assumption, there are no arrivals into station 2 whenever $x_{k2}\geqslant b$; and when there are arrivals, these occur in a Poisson stream with rate a. Hence, the process x_{k2}, $k=0,1,\cdots$ is stochastically dominated by the queue size process of an $M/M/1/b$ queue with arrival rate a and service rate μ. Let $\overline{K}_2$ be the expected time (measured in number of transitions) to empty this queue if it initially starts with b customers. Clearly, $E(K_2)<\overline{K}_2$. On the other hand, since x_{k1} can decrease by at most one per step, therefore, $EK_1\geqslant x_1\wedge n$. Hence, $\delta\leqslant c_2\overline{K}_2-c_1(x_1\wedge n)<0$ for x_1, n large. From (5.1) and (5.2), we conclude that

$$V_n(Tx)-V_n(x)=U_n(x)<0 \tag{5.3}$$

for $x=(x_1,b)$ and x_1,n large. But if (5.3) holds, then $S_n(x_1)>b$, contradicting the assumption. □

Let P denote the set of all policies over the infinite horizon. For $\pi\in P$, let

$$V_n(\pi,x)=E_x\sum_0^{n-1}c'x_k^\pi \tag{5.4}$$

where x_k^π, $k=0,1,\cdots$ is the state process corresponding to π and x is the initial state. Then

$$V_n(x)=\min_P V_n(\pi,x). \tag{5.5}$$

Let

$$\overline{V}(\pi,x)=\overline{\lim_n}\frac{1}{n}V_n(\pi,x), \tag{5.6}$$

$$\overline{V}(x)=\inf_P\overline{V}(\pi,x). \tag{5.7}$$

From (5.5) and (5.6), $\overline{V}(\pi,x)\geqslant\overline{\lim}(1/n)V_n(x)$, and so

$$\overline{V}(x)\geqslant\overline{\lim}\frac{1}{n}V_n(x). \tag{5.8}$$

Recall that the switching curves $S_n(x_1)$ are increasing in n, and so the following limit exists:

$$S(x_1)=\lim_n S_n(x_1).$$

Abusing notation slightly, let S denote the stationary policy defined by the switching curve S. Our objective is to show that: 1) under S, the Markov process $\{x_k^S\}$ is ergodic and so $\overline{V}(S,x)=\overline{V}(S)$ is independent of x, and 2) $\overline{V}(S)=\overline{\lim}(1/n)V_n(x)$ so that S minimizes the long-run average cost. It is assumed henceforth that $\lambda < \min(a, \mu)$.

Let σ denote the stationary policy under which $z_k \equiv 1$. In this case, the network consists of two $M/M/1$ queues in tandem with arrival λ and service rates a and μ in stations 1 and 2, respectively. Since $\lambda < a \wedge \mu$, therefore $\{x_k^\sigma\}$ is ergodic and $\bar{V}(\sigma, x) = \bar{V}(\sigma) < \infty$ is independent of x. From (5.7) and (5.8), we get

$$\overline{\lim} \frac{1}{n} V_n(x) \leq \bar{V}(\sigma). \tag{5.9}$$

Lemma 5.2: The Markov process $\{x_k^S\}$ is ergodic.

Proof: Let $|x| = x_1 + x_2$ denote the total number of customers in the system. Let $\epsilon > 0$. By (5.9), for all n sufficiently large,

$$\frac{1}{n} V_n(x) = \frac{1}{n} \sum^{n-1} E_x c' x_k^{\sigma^n} \leq \bar{V}(\sigma) + \epsilon \tag{5.10}$$

where σ^n is the policy which achieves the minimum in (5.5). Since $0 < c_1 \leq c_2$, (5.10) implies that

$$\frac{1}{n} \sum^{n-1} E_x |x_k^{\sigma^n}| \leq \frac{1}{c_1} [\bar{V}(\sigma) + \epsilon].$$

The policy σ^n is defined by the switching curve S_k when there are k steps to go, and $S_k \leq S$. Hence, if $x_k^S = x_k^{\sigma^n}$ for some k, then the service rate at station 1 under S is always greater than under σ^n. Considerations based on stochastic dominance then lead to

$$E_x |x_k^S| \leq E_x |x_k^{\sigma^n}|,$$

and so

$$\frac{1}{n} \sum^{n-1} E_x |x_k^S| \leq \frac{1}{c_1} [\bar{V}(\sigma) + \epsilon].$$

Hence,

$$\overline{\lim} \frac{1}{n} \sum^{n-1} E_x |x_k^S| \leq \frac{\bar{V}(\sigma)}{c_1}, \tag{5.11}$$

which clearly implies ergodicity. □

The following consequence of (5.11) will be useful later.

Corollary 5.1: For any $\epsilon > 0$, there is a finite set of states $A \subset N^2$ such that

$$\overline{\lim} \frac{1}{n} \sum^{n-1} c' x_k^S 1(x_k^S \notin A) \leq \epsilon. \tag{5.12}$$

Corollary 5.2: For $(m_1, m_2) \in N^2$, define

$$A(m_1, m_2) = \{x | x_1 \leq m_1 - 1 \text{ and } x_2 \leq S(x_1) \wedge (m_2 - 1)\}.$$

Then for any $\epsilon > 0$, there exists (m_1, m_2) such that $A(m_1, m_2)$ satisfies (5.12).

Proof: This follows from Corollary (5.1) and the fact that the set $\{x | x_1 \geq S(x_2) + 1\}$ is transient under the policy S. □

By Lemma 5.2, $\bar{V}(S, x) = \bar{V}(S)$ is independent of x. To show that $\bar{V}(S) = \overline{\lim}(1/n) V_n(x)$, we compare $V_n(S, x)$ and $V_n(x) = V_n(\sigma^n, x)$. The next lemma will be used in estimating $|V_n(S, x) - V_n(\sigma^n, x)|$.

Lemma 5.3: Let $V_k(\cdot)$, $k = 0, 1, \cdots$ be a sequence of functions and let x_k, $k = 0, 1, \cdots$ be a process with values in N^2 such that

$$E\{V_k(x_k) - V_{k+1}(x_{k+1}) | x_k = x\} = c'x.$$

Let K be a stopping time of the process $\{x_k\}$ and let Y_K be an $\underline{F}_K^x$-measurable variable such that $V_K(Y_K) \leq V_K(x_K)$ a.s. Let $\{y_k\}$ be another process such that

$$y_k = x_k, \qquad k \leq K - 1$$
$$y_K = Y_K,$$
$$E\{V_k(y_k) - V_{k+1}(y_{k+1}) | y_k = y, k > K\} = c'y.$$

Then

$$E_x \sum^{n-1} c' x_k \geq E_x \sum^{n-1} c' y_k.$$

(Here $\underline{F}_k^x$ is the σ-field generated by $x_0, \cdots, x_k$).

Proof:

$$\begin{aligned} E_x \sum^{n-1} c' y_k &= E_x \left\{ \sum^{K-1} c' y_k + V_K(Y_K) \right\} \\ &= E_x \left\{ \sum^{K-1} c' x_k + V_K(Y_K) \right\} \\ &\leq E_x \left\{ \sum^{K-1} c' x_k + V_K(x_K) \right\} \\ &= E_x \sum^{n-1} c' x_k. \end{aligned}$$

□

Lemma 5.4: $\bar{V}(S) = \overline{\lim}(1/n) V_n(x)$.

Proof: The intuitive idea of the proof is simple. Since $\{x_k^S\}$ is ergodic, it will stay in a bounded set A for most of the time, during which the control will be the same as that given by S_n for large n. An additional estimate bounds the cost when $x_k^S \notin A$. Now we give the details. Fix an initial state x. Fix $\epsilon > 0$ and select $A = A(m_1, m_2)$ as in Corollary 5.2. Since the switching curves are integer valued, there exists m such that (see Fig. 3)

$$S_n(x_1) \wedge m_2 = S(x_1) \wedge m_2, \qquad \text{for } x_1 \leq m_1 \text{ and } n > m.$$

Select $n > m$ such that

$$\frac{1}{n} E_x \sum_{n-m+1}^{n-1} c' x_k^S < \epsilon, \tag{5.13}$$

$$\frac{1}{n} E_x \sum_0^{n-1} c' x_k^S 1(x_k^S \notin A) < \epsilon. \tag{5.14}$$

Define the return times to A by $K_0 \equiv 0$:

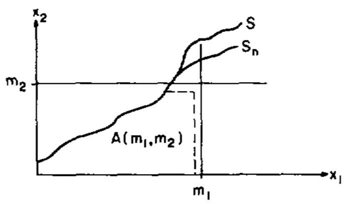

Fig. 3. Illustration for Lemma 5.4.

$$K_r = \min\{k > K_{r-1} \,|\, x_k^S \in A,\, x_{k-1}^S \notin A\} \wedge n, \qquad r \geq 1.$$

For $r \geq 1$, let

$$y_k^r = \begin{cases} x_k^S, & k \leq K_{r-1} \\ \phi\left(K_r, k; x_{K_r}^S\right), & k \geq K_r \end{cases}$$

where $\phi(k_1, k_2, x)$ is the state process at time k_2 if the state at time k_1 is x and policy σ^n is used, i.e., $\phi(k_2, k_1, x) = x_{k_2}^{\sigma^n}$ given that $x_{k_1}^{\sigma^n} = x$.

For $r = 0$, let $w_k^0 = x_k^{\sigma^n}$, and for $r \geq 1$, let

$$w_k^r = \begin{cases} w_k^{r-1}, & k \leq K_r - 1 \\ \phi\left(K_r, k; x_{K_r}^S\right), & k \geq K_r. \end{cases}$$

We claim first of all that

$$E_x \sum_0^{n-m} c' w_k^r \leq E_x \sum_0^{n-m} c' w_k^{r-1}. \tag{5.15}$$

Indeed, observe that on $\{K_r \leq n - m\}$,

$$|x_{K_r}^S| \leq |w_{K_r}^{r-1}| \quad \text{and} \quad x_{K,2}^S \geq w_{K,2}^{r-1}.$$

This implies that

$$V_{n-K_r}\left(x_{K_r}^S\right) \leq V_{n-K_r}\left(w_{K_r}^{r-1}\right) \tag{5.16}$$

since $x_{K_r}^S$ lies "below" the switching curve S_{n-K_r}, and since in this region $V_{n-K_r}(x_{K_r}^S + (l, -l) + (l', 0))$ is increasing in l and l'. Use of (5.16) in Lemma 5.3 gives (5.15).

Second, by definition of y_k^r,

$$0 \leq E_x \sum_0^{n-m} c' y_k^r - E_x \sum_0^{n-m} c' w_k^r \leq E_x \sum_0^{K_r} c' x_k^S 1\left(x_k^S \notin A\right);$$

and so, using (5.15),

$$E_x \sum_0^{n-m} c' y_k^r \leq E_x \sum_0^{n-m} c' w_k^0 + E_x \sum_0^{K_r} c' x_k^S 1\left(x_k^S \notin A\right) \leq V_n(x) + n\epsilon \tag{5.17}$$

since $w_k^0 = x_k^{\sigma^n}$ and by (5.14).

Finally, since $y_k^r \to x_k^S$ as $r \to \infty$, (5.17) implies that

$$E_x \sum_0^{n-m} c' x_k^S \leq V_n(x) + n\epsilon,$$

which, together with (5.13), gives

$$\bar{V}(S) \leq \overline{\lim}\, \frac{1}{n} V_n(x) + 2\epsilon. \qquad \square$$

Lemmas 5.2 and 5.4 give the main result.

Theorem 5.1: The stationary policy defined by the switching curve S minimizes the long-run average cost. Moreover, the resulting Markov process is ergodic.

VI. Conclusions

We have shown that the optimal policy which minimizes the cost over the infinite horizon is "bang-bang" and characterized by a monotonically increasing switching curve $x_2 = S^\beta(x_1)$ were $\beta \leq 1$ is the discount factor. The result is intuitively obvious: the "bang-bang" nature is a consequence of the fact that there is no service cost, and the monotonicity of the switching curve can be anticipated once it is surmised that the difference in the total cost incurred by adding a marginal customer to the second queue rather than adding him to the first queue must increase with the number of customers in the second queue. This intuition also suggests the form of the optimal policy in more general networks. However, substantiation of this intuition even in the simple case treated here is a nontrivial exercise, and differs from the case of a single queue in two important respects. First, the optimal processes for the discounted problem $\beta < 1$ may not be ergodic, so that the average cost case cannot be studied by taking a limit as $\beta \to 1$. Second, the convexity of the value function, which is critical for further analysis and which is implicit in the intuitive argument given above, is difficult to establish. The argument for convexity put forward here extends to more complex networks.

Note Added in Proof: V. Borkar, B. Hajek, and S. Stidham have recently and independently announced generalizations of Theorem 3.1 to the case of networks of queues.

References

[1] R. Doshi, "Optimal control of the service rate in an $M/G/1$ queuing system," *Adv. Appl. Prob.*, vol. 10, pp. 682–701, 1978.

[2] E. Gallisch, "On monotone optimal policies in a queuing model of $M/G/1$ type with controllable service time distribution," *Adv. Appl. Prob.*, vol. 11, pp. 870–887, 1979.

[3] T. Grabill, "Optimal control of service facility with variable exponential service time and constant arrival rate," *Management Sci.*, vol. 18, pp. 560–566, 1972.

[4] T. B. Grabill, D. Gross, and M. J. Magazine, "A classified bibliography of research on optimal design and control of queues," *Oper. Res.*, vol. 25, no. 2, pp. 219–232, 1977.

[5] S. A. Lippman, "Semi-Markov decision processes with unbounded rewards," *Management Sci.*, vol. 19, no. 7, pp. 717–731, 1973.

[6] ——, "Applying a new device in the optimization of exponential queuing systems," *Oper. Res.*, vol. 23, no. 4, pp. 687–710, 1978.

[7] T. Mitchell, "Optimal service-rate selection in an $M/G/1$ queue," *SIAM J. Appl. Math.*, vol. 24, pp. 19–35, 1973.

[8] R. Schassberger, "A note on optimal service selection in a single server queue," *Management Sci.*, vol. 21, pp. 1326–1331, 1975.

[9] H. Tijms, "On the optimality of a switch over policy for controlling the queue size in a $M/G/1$ queue with variable service rate," Mathematisch Centrum, Amsterdam, The Netherlands, Rep. BW 25/75, 1975.

[10] H. C. Tijms and F. A. Van der Duyn Schouten, "Inventory control

with two switch-over levels for a class of $M/G/1$ queuing systems with variable arrival and service rate," *Stochastic Processes and Their Appl.*, vol. 6, pp. 213–222, 1978.
[11] P. Varaiya, *Notes on Optimization*. New York: Van Nostrand Reinhold, 1972.

Zvi Rosberg was born in Germany on July 25, 1947. He received the B.Sc., M.A., and Ph.D. degrees from the Hebrew University, Jerusalem, Israel, in 1971, 1974, and 1978, respectively.

From 1972 to 1978 he was a Senior System Analyst in the General Computer Bureau of the Israeli government. From 1978 to 1979 he had a Research Fellowship at the Center of Operation Research and Econometric (CORE), Catholic University of Louvain, Louvain, Belgium. From 1979 to 1980 he was a Visiting Assistant Professor at the University of Illinois, Urbana. At present, he holds the position of Lecturer in the Faculty of Computer Science, Technion—Israel Institute of Technology, Haifa.

Pravin P. Varaiya (M'68–SM'78–F'80), for a photograph and biography, see this issue, p. 655.

Jean C. Walrand (S'71–M'74) was born in Belgium in 1951. He received the Ingenieur degree in electronics from the Université de Liege, Liege, Belgium, in 1974, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1979.

He was with the School of Electrical Engineering, Cornell University, Ithaca, NY, from 1979 to 1981. He is now with the Department of Electrical Engineering and Computer Science, University of California, Berkeley. His research interests are in communication networks, stochastic control, and decentralized systems.

Distributed Dynamic Programming

DIMITRI P. BERTSEKAS, SENIOR MEMBER, IEEE

***Abstract*—We consider distributed algorithms for solving dynamic programming problems whereby several processors participate simultaneously in the computation while maintaining coordination by information exchange via communication links. A model of asynchronous distributed computation is developed which requires very weak assumptions on the ordering of computations, the timing of information exchange, the amount of local information needed at each computation node, and the initial conditions for the algorithm. The class of problems considered is very broad and includes shortest path problems, and finite and infinite horizon stochastic optimal control problems. When specialized to a shortest path problem the algorithm reduces to the algorithm originally implemented for routing of messages in the ARPANET.**

RECENT advances in microcomputer technology have intensified interest in distributed computation schemes. Aside from modular expandability, other potential advantages of such schemes are a reduction in computation time for solving a given problem due to parallelism of computation, and elimination of the need to communicate problem data available at geographically dispersed data collection points to a computation center. The first advantage is of crucial importance in real-time applications where problem solution time can be an implementation bottleneck. The second advantage manifests itself for example in applications involving communication networks where there is a natural decentralization of problem data acquisition.

Manuscript received January 13, 1981; revised June 29, 1981. Paper recommended by W. E. Schmitendorf, Chairman of the Optimal Systems Committee. This work was supported in part by the National Science Foundation under Grant NSF/ECS 79-19880.

The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

The structure of dynamic programming naturally lends itself well to distributed computation since it involves calculations that to a great extent can be carried out in parallel. In fact it is trivial to devise simple schemes taking advantage of this structure whereby the calculation involved in each iteration of the standard form of the algorithm is simply shared by several processors. Such schemes require a certain degree of synchronization in that all processors must complete their assigned portion of the computation before a new iteration can begin. As a result complex protocols for algorithm initiation and processor synchronization may be necessary, and the speed of computation is limited to that of the slowest processor. These drawbacks motivate distributed algorithms whereby computation is performed asynchronously at various nodes and independently of the progress in other nodes. Their potential advantages are simpler implementation, faster convergence to a solution and, possibly, a reduction in information exchange between computation nodes.

This paper considers an asynchronous distributed algorithm for a broad class of dynamic programming problems. This class is described in Section II. The distributed computation model is described in Section III. It is shown in Section IV that the algorithm converges to the correct solution under very weak assumptions. For some classes of problems convergence in finite time is demonstrated. These include shortest path problems for which the distributed algorithm of this paper turns out to be essentially the same as the routing algorithm originally implemented in the ARPANET in 1969 [1]. To our knowledge there is no published proof of convergence of this algorithm.

0018-9286/82/0600-0610$00.75